PATTERNS OF MOLECULAR VARIATION IN PLANT POPULATIONS

R. W. ALLARD and A. L. KAHLER
UNIVERSITY OF CALIFORNIA, DAVIS

1. Introduction

It has recently been argued [8], [9] that rate of evolution at the molecular level is greater than can be accounted for by natural selection and hence that a large part of observed molecular changes must be selectively neutral. This discussion will be concerned with some experimental results, from studies of a number of different species of plants, that bear on this question. These results will be illustrated here in terms of several representative examples taken from studies of various general enzyme systems in two of the plant species under study, cultivated barley (*Hordeum vulgare*), and the Slender Wild Oat (*Avena barbata*).

The electrophoretic procedures followed are standard ones [6], [11] and consequently they need not be described here. In applying the electrophoretic techniques, we have adopted the procedure of working out the formal genetics of all bands that appear at different migrational distances for five or more enzyme systems in each species chosen for study. An example of such a full analysis of banding patterns is given in Figure 1. This figure shows, in schematic form, some banding patterns observed in a worldwide survey of esterases in cultivated barley and its wild ancestor. Bands appear in seven zones, designated A through G.

Formal genetic studies show that banding patterns for differences in migrational distances are governed by a single locus in each zone, or by seven loci in all. In the A and B zones homozygotes are single banded and heterozygotes double banded. In the C and D zones, homozygotes are double banded and heterozygotes are quadruple banded (triple banded when leading and trailing bands for two alleles are in juxtaposition). Null alleles (no band), which are recessive to alleles which produce bands, are found at the B, D, E, F and G loci. Loci A, B, and C are very tightly linked, as shown in Figure 2. Locus A is located between loci B and C, 0.0023 to the right of B and 0.0048 to the left of C. Inheritance of banding patterns for five other enzyme systems (phosphatase, leucine aminopeptidase, peroxidase, amylase, and malate dehydrogenase) which have been studied in barley also usually feature codominance.

Supported in part by grants from the National Institutes of Health (GM-10476) and from the National Science Foundation (GB-13213).

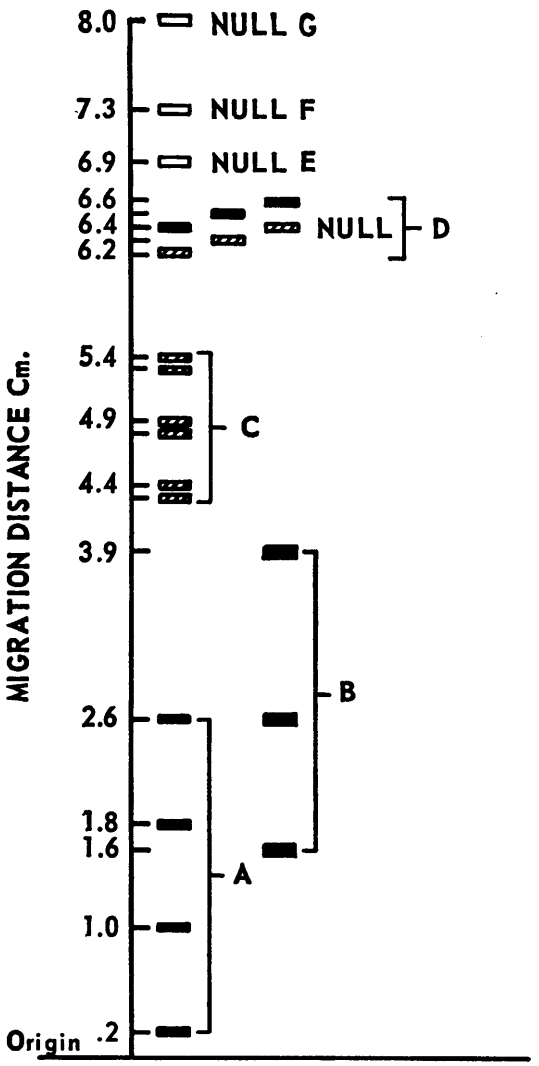

FIGURE 1

Schematic representation showing the migrational distances of some esterase electrophoretic variants observed in a worldwide survey of cultivated barley and its wild ancestor [6].

Formal genetic studies of all bands observed for these six enzyme systems establish that they are governed by 17 loci in total. This assumes that one invariant phosphatase band which has appeared uniformly in a worldwide sample of barley represents a single locus which is fixed for one allele. Since this band has appeared in all of the more than 400,000 barley plants which have been

examined, either mutation rate is unusually low at this locus, or mutations which affect migrational distance are lethal, that is, selection is very strong at this locus. The remaining 16 loci have from two up to ten or more allelic forms. This particular sample of enzymes therefore indicates that about 93 per cent of loci are not only capable of mutating to allelic forms affecting migrational distance, but also that the mutant forms can and have become established in populations. In a mutation rate study of five polymorphic loci in barley, more than 68,000 individuals, representing more than 680,000 possible mutational events, have now been examined. Since no mutants have yet been found, it

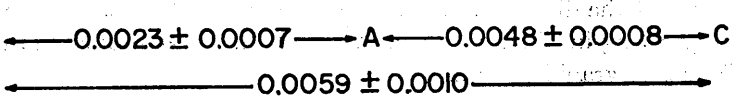

FIGURE 2

Linkage relationships among esterase loci *A*, *B*, and *C* in cultivated barley [6].

appears that the rate of mutation for migrationally detectable amino acid substitutions in these enzyme molecules is probably not higher than $1/10^5$ and hence that loci governing these variants are not unusually mutable.

2. Patterns of geographical variability in barley

A study of enzymatic variation in about 1500 entries in the world collection of barley maintained by the U.S. Department of Agriculture has shown that different populations from within any single local area often differ sharply in allelic frequencies. Allelic frequencies also differ from one ecological situation to another within a limited geographical area. Over longer distances clines are discernible, up to and including clines of continental proportions. Table I gives

TABLE I

RELATIVE ALLELIC FREQUENCIES (WITHIN THREE MAJOR CONTINENTAL AREAS) IN THE ESTERASE *B* LOCUS IN BARLEY [7]

Alleles	Relative allelic frequency		
	Europe	Middle South Asia	Orient
$B^{1.6}$	.06	.01	.00
$B^{2.0}$	.02	.03	.02
$B^{2.7}$	.86	.68	.37
$B^{3.9}$	.05	.10	.29
B^{N}	.01	.18	.32

an example of such a cline. This table shows the relative frequencies of five different alleles of the esterase *B* locus in Europe, Middle South Asia and the Orient. The slow migrating $B^{1.6}$ allele occurs in frequency of six per cent in

Europe but falls off in frequency to the east until it is very rare in the Orient. The $B^{2.7}$ allele, which is frequent in Europe, also falls off in frequency to the east. The $B^{3.9}$ and B null alleles, on the other hand, are rare in Europe and increase in frequency to the east. Extensive differentiation in allelic frequencies is thus the rule in barley on both macro- and microgeographical scales. It is equally the rule in various natural populations of plants we have studied, with one significant exception which will be discussed later. This extensive differentiation in allelic frequencies between different populations of plants is in accord with predictions based on the drift of neutral alleles [8]. However, it is also in accord with the proposition that the differentiation is adaptive and that it reflects the effects of selection operating in different ways in the different environments associated with different geographical areas.

3. Barley Composite Crosses II and V

3.1. *Description.* In discussing experiments which bear on the factors that are responsible for these observed patterns of variability in plants, we will focus attention on two experimental populations of barley, Composite Crosses II and V, and use them as a base line from which to make comparisons with various natural populations. Composite Cross V (CCV) was developed from intercrosses among 30 varieties of barley representing the major barley growing areas of the world [13]. In 1937 the 30 parents were crossed in pairs and during the next three years the F_1 hybrids of each cycle were again paircrossed to produce a single hybrid stock involving all 30 parents. This hybrid stock was then allowed to reproduce by its natural mating system, which in barley is one of about 99 per cent self fertilization and about 1 per cent outcrossing. The initial selfed generation, designated F_2, was grown in 1941 and the F_3 and all subsequent generations were grown from random samples of seeds taken from the harvest of the preceding generation. The plot was managed according to normal agricultural practice with no conscious selection practiced at any time. Viable seed has been maintained by keeping part of the harvest of each generation in storage and growing these reserve seeds at about ten generation intervals. The F_4 is the earliest generation for which viable seed of CCV is available at present.

Composite Cross II (CCII) is a substantially older experimental population than CCV and it also differs from CCV in parentage and method of synthesis. CCII was developed in 1929 by pooling equal amounts of F_1 hybrid seed from the 378 intercrosses among its 28 parents. The management of the two populations since their syntheses has, however, been the same. It should be noted that, in the 28 years that CCV and 42 years that CCII have been grown at Davis, California, temperature, rainfall, and many other factors of the environment have fluctuated sharply from year to year, and that they have also fluctuated in longer cycles.

3.2. *Parents.* Each of the parents of Composite Crosses II and V is an entry in the world barley collection maintained by U.S. Department of Agriculture.

The introduction of items into this collection is on the basis of a small sample of seeds. Thereafter each entry is maintained by growing a short row (from seed of on-type plants) whenever the seed supply is nearly exhausted. Consequently, entries in this collection are subject to severe founder effect at the time of their introduction into the collection, and also to recurring drastic restriction in population size (often to $N < 10$) thereafter. Furthermore, since barley is nearly completely self fertilized, this imposes an additional restriction on effective population size. Drift is therefore expected to be an extremely powerful force within each entry in the world collection. Consequently, it can be predicted that any polymorphism, original or new, involving *neutral alleles* will be transient within entries of the world collection and that fixation for one allele will occur at any polymorphic locus in a very few generations. In other words, the entries in the world collection are expected to be homogeneous and homozygous pure lines.

To determine whether this is the case, the genotype of each of the 28 entries in the world collection which were parents of CCII and 30 entries which were parents of CCV was determined by assaying 100 or more individuals within each entry. The sort of result obtained can be illustrated with the sample of data given

TABLE II

ALLELIC VARIABILITY WITHIN THE 28 PARENTS OF BARLEY COMPOSITE CROSS II [1]

Locus	Alleles	Monomorphic parents	Polymorphic parents		
			2 Alleles	3 Alleles	4 Alleles
A	0.2, 1.0, 1.8	6/28	15/28	7/28	
B	1.6, 2.0, 2.7, 3.0, 3.9, *N*	17/28	7/28	3/28	1/28
C	4.4, 4.9, 5.4	10/28	11/28	7/28	
D	6.2, 6.4, 6.5, 6.6, *N*	4/28	12/28	11/28	1/28

in Table II, which summarizes allelic variability for Esterase loci *A*, *B*, *C* and *D* in the parents of CCII.

Three alleles of the esterase *A* locus were represented in the 28 parents. Six of the 28 parents were monomorphic or fixed for one or the other of these three alleles. However, 15 of the parents were polymorphic with two alleles present and seven were polymorphic with three alleles present. At the *B* locus, six different alleles were represented in the parents. Seventeen of the 28 parents were monomorphic at this locus. The majority were fixed for the $B^{2.7}$ allele, which is expected because, as we saw earlier, this allele is very frequent on a worldwide basis. The 11 remaining parents were polymorphic for two, three, or four alleles at the *B* locus. The majority of the parents were also polymorphic for the *C* and *D* loci, including several which were polymorphic for three or four alleles. Considering all four loci simultaneously, only two of the 28 parents were monomorphic at all four loci, whereas three were polymorphic at one locus, five were polymorphic at two loci, nine were polymorphic at three loci and nine were

polymorphic at all four loci. When more loci were included, no parent was entirely monomorphic.

Results are similar with the parents of CCV, and with a large number of other entries in the world barley collection which have been assayed electrophoretically. It is therefore apparent the entries in the world collection of barley are extensively polymorphic. In view of the high degree of self fertilization, and the long history of propagation of these entries in very small populations, this result is not consistent with adaptive neutrality. Adaptively neutral alleles are expected to become fixed very rapidly in such small populations. The extensive polymorphism observed, however, is consistent with certain types of balancing selection to be considered later.

3.3. *Changes in allelic frequencies.* CCII and CCV, in contrast to their parents, have been carried in very large populations in each generation. In populations of many thousands of individuals per generation, almost no drift will occur and *neutral* alleles are expected to remain constant in frequency, aside from changes due to mutation and migration. It can be stated with considerable confidence that in CCII and CCV neither mutation nor migration are factors of any consequence. Mutation can be eliminated on two counts: first, the mutation rate study mentioned earlier shows that mutation rates at the loci studied are too low to have much effect on the short term dynamics of the population; second, even though these loci are known from the study of worldwide variability in barley to be capable of mutating to many allelic forms, no alleles not present originally were found in any generation. Migration can be eliminated on this same basis; no alleles not present originally were found and such alleles are unlikely to have escaped detection had migration from the outside occurred into either CCII or CCV. Hence, if the molecular variants in these two populations are adaptively neutral, no change in allelic frequencies is expected in very large populations such as CCII and CCV. The observation that allelic frequencies remain constant in these populations, which have been grown in an environment which has fluctuated over generations, would therefore provide evidence in support of the proposition that the molecular variants are neutral. Conversely, the observation that gene frequencies change would provide evidence that they are affected by selection (or some yet undiscovered evolutionary factor producing effects parallel to those of selection).

Table III illustrates in terms of the esterase *C* locus, the sort of result that is obtained when allelic frequencies are monitored over generations in CCV. The parents contributed three alleles at this locus in CCV and the frequencies of these alleles, as inferred from their frequencies in the 30 parents, are 14.4 per cent for allele $C^{4.4}$, 24.7 per cent for allele $C^{4.9}$ and 60.9 per cent for allele $C^{5.4}$. The number of individuals assayed electrophoretically was large (from about 1000 up to nearly 4400 individuals) in each of the three early generations (4, 5, 6), four intermediate (14, 15, 16, 17), and three late (24, 25, 26) generations that were monitored. Standard errors for allelic frequencies are small (<0.01) so that changes of 0.02 in allelic frequency, or smaller, are significant. In several cases,

TABLE III

RELATIVE ALLELIC FREQUENCIES AT THE ESTERASE *C* LOCUS IN BARLEY COMPOSITE CROSS V [2]

The allelic frequencies for the initial generation are inferred from those of the 30 parents. Standard errors are < 0.01.

Generation	Number assayed	Allele $C^{4.4}$	$C^{4.9}$	$C^{5.4}$
Initial	4569	.144	.247	.609
4	1234	.033	.265	.702
5	1486	.049	.301	.650
6	1006	.074	.287	.639
14	1651	.034	.203	.763
15	2843	.082	.281	.637
16	2369	.050	.326	.624
17	2461	.077	.307	.616
24	4397	.102	.316	.582
25	3967	.211	.308	.481
26	3083	.279	.254	.467

for example, alleles $C^{4.4}$ and $C^{5.4}$ in transition from generation 24 to 25, changes in allelic frequency >0.10, that is, more than ten standard deviations, or larger occurred. Such changes are consistent with selection operating differentially in the drastically different environmental conditions to which this population was exposed in certain successive years. In a population of this size, they are not consistent with steps in a random walk by neutral alleles.

Table IV gives data for the same locus in CCII. Again significant changes in

TABLE IV

RELATIVE ALLELIC FREQUENCIES AT THE ESTERASE *C* LOCUS IN BARLEY COMPOSITE CROSS II [1]

The allelic frequencies for the initial generation are inferred from those of the 28 parents. Standard errors are < 0.01.

Generation	Number assayed	Allele $C^{4.4}$	$C^{4.9}$	$C^{5.4}$
Initial	3248	.112	.428	.460
7	1046	.193	.205	.602
8	1140	.190	.202	.608
9	948	.218	.165	.617
17	2398	.307	.275	.418
18	2094	.299	.282	.419
19	1903	.302	.249	.449
39	3472	.144	.010	.846
40	3075	.137	.015	.849
41	2868	.119	.018	.863

allelic frequencies occurred in certain single generation transitions, for example, for allele $C^{4.9}$ in transition from generation 8 to 9, and from generation 18 to 19. In addition, it is clear that longer term changes have also occurred. In the eight generation interval from the earlier generations (7, 8, 9) to the intermediate generations (17, 18, 19) alleles $C^{4.4}$ and $C^{4.9}$ both increased in frequency by about 0.10 at the expense of a loss in frequency of about 0.20 for allele $C^{5.4}$. In the generation interval between the intermediate generations (17, 18, 19) and the late generations (39, 40, 41) the trend reversed and allele $C^{5.4}$ gained 0.40 in frequency at the expense of allele $C^{4.4}$ and particularly at the expense of allele $C^{4.9}$, which was reduced to very low frequency in this population. Selection has not followed the same course for the *C* locus in the two populations. This is not surprising considering the different years in which the populations were grown and the different genetic backgrounds of the two populations.

Another feature of genetic change in Composite Crosses II and V is shown in Tables V and VI, which give the observed percentage of heterozygotes for three

TABLE V

PERCENTAGE OF HETEROZYGOTES FOR ESTERASE LOCI *A*, *B*, AND *C* IN BARLEY COMPOSITE CROSS II [1]

Generation	Locus *A*	*B*	*C*
7	5.36	1.05	6.31
8	3.67	0.78	2.28
9	3.37	1.26	2.64
17	2.21	0.33	2.71
18	1.39	0.77	2.01
19	0.96	0.21	1.95
39	1.44	0.69	1.10
40	0.98	0.56	0.94
41	0.49	0.14	0.83

representative loci in some early, intermediate and late generations. It can be deduced from the genotypes of the parents, and the sequence in which they were hybridized, that the F_1 generations of both CCII and V were highly heterozygous. In populations in which the mating system features more than 99 per cent of self fertilization, it is expected that about half of the initial heterozygosity will be lost per generation until an equilibrium level is approached in the fifth or sixth generation. The results given in Tables V and VI show that heterozygosity had been reduced to low levels in the earliest generations of CCII (generation 7) and CCV (generation 4) available for study and that no further consistent change occurred in the later generations. Thus, the pattern of change followed expected patterns, at least in a general way, in both populations.

The question which must now be asked is whether the observed changes in

TABLE VI

PERCENTAGE OF HETEROZYGOTES FOR ESTERASE LOCI *A*, *B*, AND *C* IN VARIOUS GENERATIONS OF BARLEY COMPOSITE CROSS V [2]

Generation	Locus *A*	*B*	*C*
4	7.12	3.83	9.24
5	4.31	1.34	5.72
6	3.29	0.00	1.89
14	2.44	1.87	7.20
15	2.78	4.37	2.33
16	1.48	1.14	3.17
17	0.65	0.08	0.61
24	2.50	2.10	7.62
25	1.61	1.06	1.48
26	1.30	0.64	1.56

genotypic frequencies fit expectations for selectively neutral alleles. One way this question can be answered is to compute theoretical inbreeding coefficients F, which assume that the relationship between gene and genotypic frequencies is solely a function of mating system. These theoretical F values can then be compared with fixation indices $\hat{F}$, which are computed from observed genotypic frequencies, and hence measure the actual relationship between gene and genotypic frequencies, that is, the inbreeding actually realized. Computation of theoretical inbreeding coefficients requires precise estimates of the proportion of selfing *versus* outcrossing. Estimates were made by assaying electrophoretically about 18,000 progeny of plants taken from ten generations of CCV and 13,000 progeny of plants from CCII. These estimates are homogeneous over loci, generations, and years within the two populations. They also agree closely with earlier estimates of outcrossing made using morphological polymorphisms, and with general experience with barley. Hence, it seems safe to use the mean observed outcrossing rate of 0.78 and 0.57 per cent for CCII and CCV, respectively, to calculate theoretical inbreeding coefficients for the two populations.

Values of the theoretical inbreeding coefficient for CCV are given in Table VII. In populations such as CCII and CCV, which were synthesized by random crossing among diverse parents, F is expected to be zero in the original generation. In subsequent generations, with more than 99 per cent self fertilization, F is expected to follow approximately the series 0, $\frac{1}{2}$, $\frac{3}{4}$, $\frac{7}{8}$ $\cdots$ until in generation 6 or 7 it is expected to approach its equilibrium value of 0.989 (see footnote, Table VII). Table VII also gives a representative sample of fixation indices computed from the observed gene and genotypic frequencies. In virtually all cases the theoretical inbreeding coefficient is larger than the fixation index, which shows that there are consistent excesses of heterozygotes over levels expected on the basis of mating system alone. Thus, this result also does not

TABLE VII

THEORETICAL INBREEDING COEFFICIENTS F AND OBSERVED FIXATION INDICES $\hat{F}$ FOR REPRESENTATIVE GENOTYPES AND GENERATIONS IN BARLEY COMPOSITE CROSS V [2]

$$F^n = \frac{s}{1+t}[1 - (\tfrac{1}{2}s)^n], \quad \hat{F} = 1 - \frac{H_{ij}}{2p_ip_j}$$

Generation	Theoretical F	Observed fixation indices				
		$A^{0.2}A^{1.8}$	$A^{1.0}A^{1.8}$	$B^{1.6}B^{2.7}$	$C^{4.4}C^{5.4}$	$C^{4.9}C^{5.4}$
4	.928	.830	.887	.779	.670	.795
5	.959	.890	.923	.830	.885	.876
14	.989	.971	.951	.903	.752	.817
15	.989	.947	.947	.978	.983	.948
25	.989	.965	.971	.970	.983	.975
26	.989	.955	.980	.983	.986	.966

conform to the hypothesis of neutral alleles. Again the simplest explanation for the excess of heterozygotes appears to be some sort of balancing selection which leads to a net advantage of heterozygotes in reproduction.

3.4. *Two locus interactions.* Before discussing expectations and results for pairs of loci considered simultaneously, it is necessary to define the two locus gametic and zygotic arrays in a population. For two loci with two alleles each (A, a and B, b) let allelic frequencies be p_1, q_1, and p_2, q_2, respectively, and let the frequencies of the four gametic types AB, Ab, aB, and ab be g_1, g_2, g_3, and g_4. There are ten possible genotypes with frequencies (f_i) as follows:

(1)

	AA	Aa	aa
BB	f_1	f_4	f_8
Bb	f_2	$f_5(AB/ab)$ $f_6(Ab/aB)$	f_9
bb	f_3	f_7	f_{10}.

Linkage equilibrium is defined as the condition in which the equilibrium frequencies of the gametic ditypes correspond to the products of the appropriate gene frequencies, that is,

$$(2) \qquad \hat{g}_1 = \hat{p}_1\hat{p}_2, \quad \hat{g}_2 = \hat{p}_1\hat{q}_2, \quad \hat{g}_3 = \hat{q}_1\hat{p}_2, \quad \hat{g}_4 = \hat{q}_1\hat{q}_2.$$

In linkage equilibrium situations, gametic and zygotic frequencies thus correspond to the products of single locus gene frequencies. For nonequilibrium situations gametic frequencies are given by:

$$(3) \quad \hat{g}_1 = \hat{p}_1\hat{p}_2 + D, \quad \hat{g}_2 = \hat{p}_1\hat{q}_2 - D, \quad \hat{g}_3 = \hat{q}_1\hat{p}_2 - D, \quad \hat{g}_4 = \hat{q}_1\hat{q}_2 + D,$$

where $D = \hat{g}_1\hat{g}_4 - \hat{g}_2\hat{g}_3$. Values of D range from -0.25 (all Ab and aB) to $+0.25$ (all AB and ab). When $D \neq 0$, gametic and zygotic frequencies do not correspond to the products of single locus gene frequencies.

The conditions necessary for the development and maintenance of linkage

disequilibrium are the simultaneous existence of epistatic selection and certain combinations of tight linkage and/or inbreeding [5], [10]. Since selection is a requirement for linkage disequilibrium, pairs of adaptively neutral alleles originally in linkage equilibrium will not develop linkage disequilibrium ($D \neq 0$), or if $D \neq 0$ for some reason (for example, sampling effects due to the limited number of parents as in CCII and V), $D \rightarrow 0$ at a rate depending on the crossover value between the loci and/or the degree of inbreeding. This implies that, for neutral alleles, two locus zygotic frequencies should be predictable from the products of single locus frequencies, and that D should go to zero.

In illustrating the observed results in Composite Crosses II and V, it is convenient to start with the observed zygotic arrays for the esterase B and C loci in various generations of CCV, (Table VIII). There are three alleles at locus

TABLE VIII

DEVIATIONS FROM PRODUCTS OF ONE LOCUS NUMBERS FOR ESTERASE LOCI B AND C IN GENERATIONS 6, 17, AND 26 OF BARLEY COMPOSITE CROSS V [14]

	Locus B	Locus C $C^{4.4}C^{4.4}$	$C^{4.9}C^{4.9}$	$C^{5.4}C^{5.4}$
Generation 6	$B^{1.6}B^{1.6}$	+17	−5	−11
$N = 1006$	$B^{2.7}B^{2.7}$	−28	+1	+27
$\chi^2 = 124.9$	$B^{3.9}B^{3.9}$	+10	+4	−15
Generation 17	$B^{1.6}B^{1.6}$	+65	−37	−28
$N = 2461$	$B^{2.7}B^{2.7}$	−130	+45	+84
$\chi^2 = 1236.5$	$B^{3.9}B^{3.9}$	+65	−8	−56
Generation 26	$B^{1.6}B^{1.6}$	+387	−133	−247
$N = 3083$	$B^{2.7}B^{2.7}$	−492	+145	+354
$\chi^2 = 2187.8$	$B^{3.9}B^{3.9}$	+107	−9	−100

B and hence six possible genotypes at this locus. The same is the case for locus C. Considering both loci simultaneously, there are 36 possible genotypes, among which only the nine homozygous combinations are shown to keep the table within acceptable size. The values in this table are deviations of two locus numbers from numbers predicted from single locus frequencies. In generation 6 there is an indication that certain combinations of alleles at the two loci interact favorably with each other (for example, $B^{1.6}$ and $C^{4.4}$) in their homozygous combinations and that others interact unfavorably (for example, $B^{2.7}$ and $C^{4.4}$). By generation 17 the deviations from marginal frequencies have become very large and by generation 26 they have become larger still. This point is brought out by χ^2 values which show rapid increase over generations. Note that each of the three alleles at locus B interacts favorably in at least one of its homozygous combinations with each of the three alleles at locus C, and that each interacts unfavorably with at least one C locus allele.

Similar epistatic interactions, both favorable and unfavorable, also occur for

the heterozygous combinations of alleles at the two loci, which are not shown here. This sort of epistatic interaction on the fitness scale leads to a balancing type of selection which can be shown to promote the development and maintenance of very stable polymorphisms at both loci. These data show in terms of the zygotic array, how the alleles at the two loci have gone from random associations in early generations, as expected in populations synthesized by random crossing between diverse parents, to very specific associations in later generations.

TABLE IX

LINKAGE DISEQUILIBRIUM (GAMETIC UNBALANCE) VALUES FOR THE TIGHTLY LINKED ESTERASE LOCI *A*, *B*, AND *C* IN BARLEY COMPOSITE CROSS V [14]

$D = g_1 \cdot g_4 - g_2 \cdot g_3$, where g_1, g_2, g_3 and g_4 are the frequencies of the gametic ditypes *AB*, *Ab*, *aB*, and *ab*, respectively.

Generation	Pairs of loci		
	A–B	*A–C*	*B–C*
5	.02	.02	.02
6	.02	.01	.03
16	.02	.05	.03
17	.02	.05	.03
25	.09	.08	.11
26	.10	.10	.11

Table IX shows the same results in terms of the gametic array for loci *BC*, and also for loci *AB* and *AC*. Initially, CCV was in near linkage equilibrium ($D = 0$) for all three pairwise combinations, but D gradually increased until it had become very large by generations 25 and 26. That such striking interaction systems have built up so rapidly implies that selection must have been of great intensity.

If the *A*, *B*, and *C* loci themselves are subject to selection, and if various alleles at these loci interact with one another in specific ways, it might be expected that the same interaction systems should develop whenever the same alleles occur together in the same population. All three alleles occur in both CCV and CCII and hence it is interesting to compare the zygotic arrays of these two populations. Deviations from expectations are very similar, in the two populations, the greatest difference being in the weaker interactions between allele $B^{3\cdot9}$ and the *C* locus alleles in CCII (Tables VIII and X). The similarity of the two zygotic arrays is brought out most clearly by comparison of generation 17 in CCV and generation 18 in CCII. Note that correspondence between the populations in these comparable generations is identical in direction, and similar in magnitude, in each combination. For the zygotic array in either CCII or CCV to have progressed from linkage equilibrium to linkage disequilibrium as a result of a random walk of neutral alleles seems unlikely. For both popula-

TABLE X

DEVIATIONS FROM PRODUCTS OF ONE LOCUS NUMBERS FOR ESTERASE LOCI *B* AND *C* IN GENERATIONS 8, 18, AND 40 OF BARLEY COMPOSITE CROSS II [1]

	Locus *B*	Locus *C* $C^{4.4}C^{4.4}$	$C^{4.9}C^{4.9}$	$C^{5.4}C^{5.4}$
Generation 8	$B^{1.6}B^{1.6}$	+55	−13	−41
N = 1140	$B^{2.7}B^{2.7}$	−116	+25	+90
χ^2 = 413.0	$B^{3.9}B^{3.9}$	+8	0	−9
Generation 18	$B^{1.6}B^{1.6}$	+96	−37	−56
N = 2095	$B^{2.7}B^{2.7}$	−170	+66	+107
χ^2 = 452.2	$B^{3.9}B^{3.9}$	+19	−5	−14
Generation 40	$B^{1.6}B^{1.6}$	+153	−2	−149
N = 3075	$B^{2.7}B^{2.7}$	−214	+3	+219
χ^2 = 1268.6	$B^{3.9}B^{3.9}$	0	0	0

TABLE XI

DEVIATIONS FROM PRODUCTS OF ONE LOCUS NUMBERS FOR THE ESTERASE LOCI *B* AND *D* (UNLINKED) IN GENERATIONS 6, 17, AND 26 OF BARLEY COMPOSITE CROSS V [14]

	Locus *B*	Locus *D* $D^{6.4}D^{6.4}$	$D^{6.5}D^{6.5}$	$D^{6.6}D^{6.6}$	$D^{N}D^{N}$
Generation 6	$B^{1.6}B^{1.6}$	−5	−1	+6	+1
N = 1006	$B^{2.7}B^{2.7}$	+4	+2	−5	−1
χ^2 = 5.1 (NS)	$B^{3.9}B^{3.9}$	+1	−1	−1	−1
Generation 17	$B^{1.6}B^{1.6}$	+19	+4	−21	−2
N = 2461	$B^{2.7}B^{2.7}$	−46	−6	+17	+19
χ^2 = 49.5	$B^{3.9}B^{3.9}$	+8	+2	+4	−17
Generation 26	$B^{1.6}B^{1.6}$	+139	−27	−77	−29
N = 3083	$B^{2.7}B^{2.7}$	−166	+29	+62	+42
χ^2 = 240.8	$B^{3.9}B^{3.9}$	−2	−3	+10	−14

tions to go from random association to the same state of organization without the guiding force of selection seems even more unlikely.

Table XI gives, in terms of esterase loci *B* and *D* in CCV, an example of the build up of a specific interaction system between two unlinked loci. In generation 6 departures from expectations based on marginal frequencies are small and not significant, that is, the alleles at the two loci occur together at random. By generation 17 there is indication that certain alleles interact favorably in their homozygous combinations with each other (for example, $B^{1.6}$ and $D^{6.4}$) and others interact unfavorably (especially $B^{2.7}$ and $D^{6.4}$). Results in generation 26 confirm the reality of the earlier trends. Table XII gives *D* values for these pairwise comparisons of unlinked loci. Again the build up of *D* shows the change from random association to an organized state. Table XIII compares the zygotic

TABLE XII

LINKAGE DISEQUILIBRIUM (GAMETIC UNBALANCE) VALUES FOR ESTERASE LOCUS *D* WITH ESTERASE LOCI *A*, *B*, AND *C* IN BARLEY COMPOSITE CROSS V [14]

$D = g_1 \cdot g_4 - g_2 \cdot g_3$, where g_1, g_2, g_3 and g_4 are the frequencies of the gametic ditypes *AB*, *Ab*, *aB*, and *ab*, respectively.

Generation	Pairs of loci		
	A–D	*B–D*	*C–D*
5	.00	.00	.02
6	.01	.00	.01
16	.01	.02	.01
17	.01	.01	.02
25	.02	.06	.04
26	.02	.05	.03

TABLE XIII

DEVIATIONS FROM ONE LOCUS NUMBERS IN COMPARABLE GENERATIONS OF BARLEY COMPOSITE CROSSES II AND V [1]

	B Locus	Locus *D*			
		$D^{6.4}D^{6.4}$	$D^{6.5}D^{6.5}$	$D^{6.6}D^{6.6}$	$D^{N}D^{N}$
Composite Cross II Generation 18	$B^{1.6}B^{1.6}$	+41	−3	−9	−30
	$B^{2.7}B^{2.7}$	−61	−5	+14	+50
	$B^{3.9}B^{3.9}$	−3	+7	−1	−2
Composite Cross V Generation 17	$B^{1.6}B^{1.6}$	+19	+4	−21	−2
	$B^{2.7}B^{2.7}$	−46	−6	+17	+19
	$B^{3.9}B^{3.9}$	+8	+2	+4	−17

arrays for the same loci and alleles in comparable generations of CCII and CCV. Again the departures from random associations are the same in direction and magnitude in the two populations.

On the basis of results such as these we conclude that genes do not exist in populations in random backgrounds. On the contrary, the normal situation is probably existence in correlated blocks within chromosomes (as with loci *A*, *B* and *C*) and also between unlinked loci (as with loci *AD*, *BD*, and *CD*). Further, natural selection operates not only on single loci but also on the correlated state. Apparently the gametic and zygotic arrays, and evolutionary changes in these arrays, cannot be described adequately in terms of gene frequencies at single loci. Descriptions of the multilocus gametic and zygotic arrays apparently have to be in terms of larger units, such as linkage blocks, whole chromosomes, or even the entire population genotype, if they are to be consonant with the observations. In other words, before we can allow conclusions about rates of

evolution based on single locus substitutions to rule out selection, we have to know much more than we do at present about interactions between loci at the level of the fitness scale.

4. Geographical variation in *Avena barbata*

Let us now turn to some observations on natural populations that are relevant to the random theory. Table XIV gives gene and genotypic frequencies for a

TABLE XIV

GENE AND GENOTYPIC FREQUENCIES IN A HILLSIDE POPULATION OF *Avena barbata* (SITE CSA) [4]

Locus	Genotype	Location 1	2	3	4	5
E_4	11	.129	.740	.734	.631	1.000
	12	.113	.016	.109	.062	.000
	22	.758	.194	.156	.308	.000
	q_2	.814	.202	.211	.338	.000
E_9	11	.564	.145	.156	.354	.000
	12	.097	.032	.078	.062	.000
	22	.339	.823	.766	.585	1.000
	q_2	.387	.839	.805	.615	1.000
APX_5	11	.190	.806	.859	.600	1.000
	12	.127	.032	.094	.062	.000
	22	.682	.161	.047	.338	.000
	q_2	.746	.177	.094	.369	.000

sample of three typical loci in a population of the Slender Wild Oat, *Avena barbata*, which occupies a site (CSA) about 200 feet wide and 400 feet long, extending up a hillside in the Coast Range near Calistoga, California. This site is mesic at the bottom of the hillside and it becomes progressively more arid up the hillside. Location 4 within the site departs from this ecological gradient in that it represents a flat area of deeper soil which is ecologically more like locations 1 and 2 at the bottom of the hillside than arid location 5 at the extreme top of the hillside.

Considering allelic frequencies at the esterase 4 and APX_5 loci, it can be seen that the faster migrating allele (allele 2) is in high frequency in the mesic location at the bottom of the hillside. Progressing up the hillside, the frequency of these alleles falls off in locations 2 and 3, increases again in location 4, and falls off to zero in location 5. The pattern for the esterase 9 locus differs only in that the slower migrating allele is in low frequency in the mesic locations and the faster migrating allele (allele 2) is in high frequency in the arid locations, becoming fixed in location 5. This progressive change in allelic frequencies, and in poly-

morphism, for these and for other loci, on this hillside in fact reflects the geographical variation which occurs throughout California [3].

In the mixed but generally mesic habitats of the Coast Range, from about Monterey northward, most populations of *A. barbata* are polymorphic with allelic frequencies falling generally within the range of those of locations 1 through 4 of site CSA. These populations are also polymorphic for many morphological characters and the extent of polymorphism for molecular and morphological traits is highly correlated [3], [12]. In the arid habitats east of the crest of the Coast Range, and in the foothills of the Sierra Nevada Mountains south to San Diego, all populations that have been analyzed are fixed for the genotype which is found in arid location 5 of the CSA site. The observation that only a single genotype occurs in the numerous isolated populations found over this very large geographical area is particularly difficult to explain by the random theory because the random theory predicts heterozygosity within populations, and it also predicts that different alleles will be present at different frequencies in different locations. In other words, it predicts the opposite of the observations.

There are also two aspects of the variability in the Coast Range that are difficult to explain by the random drift of neutral alleles. First, there is the observation that sharp geographical divergence, correlated with the details of the habitat, are maintained over very short distances (such as within site CSA), even though some pollen and some seed migration occurs between such locations. Even a little migration would homogenize allelic frequencies among populations if the molecular variants were neutral. Second, is the observation that the level of heterozygosity at all loci examined in all populations is higher than can be explained on the basis of mating system alone. An example is given in Table XV

TABLE XV

OUTCROSSING VALUES t, INBREEDING COEFFICIENTS F AND FIXATION INDICES $\hat{F}$ WITHIN SITE CSA [4]

Item	Location 1	2	3	4
Outcrossing Rate t	.027	.003	.013	.023
Inbreeding Coefficient F	.947	.994	.974	.955
	$\hat{F}$	$\hat{F}$	$\hat{F}$	$\hat{F}$
E_4	.697	.950	.673	.861
E_9	.796	.882	.752	.869
APX_5	.665	.890	.448	.867

which shows theoretical inbreeding coefficients and fixation indices for the four polymorphic locations within the CSA site. The fixation indices are substantially lower than theoretical inbreeding coefficients in all four locations. Table XVI expresses these excesses in terms of the selection that is necessary to maintain

TABLE XVI

SELECTIVE VALUES OF HOMOZYGOTES (ALLELE $ii = x$, ALLELE $jj = y$) RELATIVE TO HETEROZYGOTES TAKEN AS UNITY [4]

Locus	Location							
	CSA-1		CSA-2		CSA-3		CSA-4	
	x	y	x	y	x	y	x	y
E^4	.469	.541	.557	.549	.516	.459	.643	.632
E^9	.599	.587	.498	.519	.489	.533	.644	.652
APX_5	.495	.545	.521	.502	.503	.346	.649	.640

them. It can be seen that homozygotes have only about half the reproductive capacity of heterozygotes. The next step in this study will be to examine large enough samples within single populations to determine whether this excess of heterozygotes is due to epistatic interactions between alleles at different loci, similar to the situation in the experimental populations of barley.

5. Summary and conclusions

This discussion of enzyme variants in plant populations can be summed up in five main points.

First, there is extensive allelic variability within entries of the world collection of barley maintained by the U.S. Department of Agriculture. This variability appears to be much in excess of amounts that can be explained in such small populations on the basis of the drift of neutral alleles. However, it is consistent with certain types of rather strong balancing selection.

Second, the changes in allelic frequencies which occur from generation to generation in Composite Crosses II and V are much too large to be explained by genetic drift. The data also show that these changes cannot be due to mutation or migration, but that they are consistent with selection operating in different ways in the different environmental conditions of different years, or groups of years.

Third, comparisons of theoretical inbreeding coefficients and fixation indices show an excess of heterozygotes in Composite Crosses II and V, and in natural populations of *A. barbata*. This result is inconsistent with neutral alleles, but it is consistent with balancing selection.

Fourth, in early generations of Composite Crosses II and V, frequencies of two locus genotypes are generally in agreement with predictions based on single locus frequencies, that is, combinations of alleles at different loci are at random. However, within a few generations nonrandom associations develop which can be identified with favorable and unfavorable epistatic interactions between specific alleles at different loci. The same associations develop in the two populations. This result is consistent with very strong selection. It is not consistent with random walks of neutral alleles.

Fifth, the existence of a single genotype over a large part of the range of *A. barbata* in California is a most difficult observation to explain by the random theory, which predicts that different alleles should drift to fixation in different places.

We have studied all alleles which produce bands at migrationally different distances in our materials. None of the alleles appear to be neutral or physiologically irrelevant. Instead, all the alleles in our sample appear to be adaptive, but their effect on fitness seems to be expressed through apparently complex interactions with alleles at other loci. Until such interactions are better understood, we conclude that calculations based on single locus substitutions should be regarded cautiously, as should generalizations concerning evolutionary rates made from such calculations.

REFERENCES

[1] R. W. ALLARD, A. L. KAHLER, and M. T. CLEGG, "The dynamics of multilocus esterase polymorphisms in an experimental plant population," in preparation.
[2] R. W. ALLARD, A. L. KAHLER, and B. S. WEIR, "The effect of selection on esterase allozymes in a barley population," *Genetics*, in press.
[3] M. T. CLEGG, "Patterns of genetic differentiation in natural populations of wild oats," Ph.D. dissertation, University of California, Davis.
[4] J. L. HAMRICK and R. W. ALLARD, "Microgeographical variation in allozyme frequencies in *Avena barbata*," in preparation.
[5] S. K. JAIN and R. W. ALLARD, "The effects of linkage, epistasis and inbreeding on population changes under selection," *Genetics*, Vol. 53 (1966), pp. 633–659.
[6] A. L. KAHLER and R. W. ALLARD, "Genetics of isozyme variants in barley. I. Esterases," *Crop Science*, Vol. 10 (1970), pp. 444–448.
[7] ———, "Worldwide patterns of differentiation of esterase allozymes in barley (*Hordeum vulgare* and *H. spontaneum*)," in preparation.
[8] M. KIMURA, "Evolutionary rate at the molecular level," *Nature*, Vol. 217 (1968), pp. 624–626.
[9] J. L. KING and T. H. JUKES, "Non-Darwinian evolution," *Science*, Vol. 164 (1969), pp. 788–798.
[10] R. C. LEWONTIN and K. KOJIMA, "The evolutionary dynamics of complex polymorphisms," *Evolution*, Vol. 14 (1960), pp. 458–472.
[11] D. R. MARSHALL and R. W. ALLARD, "The genetics of electrophoretic variants in Avena," *J. Heredity*, Vol. 60 (1969), pp. 17–19.
[12] ———, "Isozyme polymorphisms in natural populations of *Avena fatua* and *A. barbata*," *Heredity*, Vol. 25 (1970), pp. 373–382.
[13] C. A. SUNESON, "An evolutionary plant breeding method," *Agronomy J.*, Vol. 48 (1956), pp. 188–190.
[14] B. S. WEIR, R. W. ALLARD, and A. L. KAHLER, "Analysis of complex allozyme polymorphisms in a barley population," *Genetics*, in press.